PERGAMON

Pattern Recognition 34 (2001) 361–373

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

A linear constrained distance-based discriminant analysis for hyperspectral image classification

Qian Du[a], Chein-I Chang[b,*]

[a]Assistant Professor of Electrical Engineering, Texas A&M University-Kingsville, Kingsville, TX 78363, USA
[b]Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA

Received 20 April 1999; received in revised form 15 October 1999; accepted 15 October 1999

Abstract

Fisher's linear discriminant analysis (LDA) is a widely used technique for pattern classification problems. It employs Fisher's ratio, ratio of between-class scatter matrix to within-class scatter matrix to derive a set of feature vectors by which high-dimensional data can be projected onto a low-dimensional feature space in the sense of maximizing class separability. This paper presents a linear constrained distance-based discriminant analysis (LCDA) that uses a criterion for optimality derived from Fisher's ratio criterion. It not only maximizes the ratio of inter-distance between classes to intra-distance within classes but also imposes a constraint that all class centers must be aligned along predetermined directions. When these desired directions are orthogonal, the resulting classifier turns out to have the same operation form as the classifier derived by the orthogonal subspace projection (OSP) approach recently developed for hyperspectral image classification. Because of that, LCDA can be viewed as a constrained version of OSP. In order to demonstrate its performance in hyperspectral image classification, Airborne Visible/InfraRed Imaging Spectrometer (AVIRIS) and HYperspectral Digital Imagery Collection Experiment (HYDICE) data are used for experiments. © 2000 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Constrained energy minimization (CEM); Linear constrained distance-based discriminant analysis (LCDA); Linear discriminant analysis (LDA); Orthogonal subspace projection (OSP); Unsupervised LCDA (ULCDA); Unsupervised CEM (UCEM); Unsupervised LDA (ULDA); Unsupervised OSP (UOSP)

1. Introduction

Remotely sensed images generally consist of a set of co-registered images taken at the same time by different spectral channels during data acquisition. Consequently, a remote sensing image is indeed an image cube with each image pixel represented by a column vector. In addition, due to a large area covered by an instantaneous field of view of remote sensing instruments, an image pixel vector generally contains more than one endmember (substance). This result in a mixture of endmembers resident within the pixel vector rather than a pure pixel considered in classical image processing. So, standard image processing techniques are not directly applicable.

Many mixed pixel classification methods have been proposed such as linear unmixing [1–8]. One of principal differences between pure and mixed pixel classification is that the former is a class membership assignment process, whereas the latter is actually endmember signature abundance estimation. An experiment-based comparative study [9] which showed that mixed pixel classification techniques generally performed better than pure pixel classification methods. Additionally, it has been also shown that if mixed pixel classification was converted to pure pixel classification, Fisher's linear discriminant analysis (LDA) [10] was among the best. It becomes obvious that directly applying pure pixel-based LDA to mixed pixel classification problems may not be

* Corresponding author. Tel.: + 1-410-455-3502; fax: + 1-410-455-3969.
E-mail address: cchang@umbc.edu (C.-I. Chang).

0031-3203/00/$20.00
PII: S0031-3203(99)00215-0

an effective way to best utilize LDA. In magnetic resonance imaging (MRI) applications [11], Soltanian-Zadeh et al. recently developed a constrained criterion to characterize brain issues for 3-D feature representation. The Soltanian-Zadeh et al.'s criterion is the ratio of the inter-distance (IED) to intra-set distance (IAD) [12] subject to a constraint that each class center must be aligned along some predeterimined directions. In order to arrive at an analytic solution, Soltanian-Zadeh et al. further made an assumption of white Gaussian noise based on which IED can be reduced to a constant. As a result, maximizing the ratio of IED to IAD is reduced to minimizing IAD. However, the white Gaussian noise assumption may not be valid in hyperspectral images since it has been demonstrated [8,13] that unknown interference such as background signatures, clutters in hyperspectral imagery were more severe and dominant than noise which may result in non-Gaussianity and nonstationarity. As a matter of fact, we will show in this paper that such assumption is not necessary and can be removed by a whitening process. By modifying Soltanian-Zadeh et al.'s approach a linear constrained distance-based discriminant analysis (LCDA) is developed in this paper for hyperspectral image classification and target detection. Two important aspects resulting from LCDA are worth being mentioned. One is to show that minimizing IAD is equivalent to minimizing the trace of the data sample covariance matrix Σ. In this case, a whitening process can be developed to decorrelate the Σ without making Gaussian assumption. Another is to show that after such a whitening processing is accomplished, LCDA can be simply carried out by orthogonal subspace projection (OSP) [1,7]. In the light of classification, LCDA can be viewed as a constrained version of OSP. As will be shown in experiments, LCDA performs significantly better than OSP.

One of advantages of LCDA is to use a constraint to steer class centers of interest along desired directions, generally orthogonal directions. Such constraint allows us to separate different classes as farther as possible. The idea of using direction constraints is not new and has been found in various applications, minimum variance distortionless response (MVDR) beamformer in array processing [14,15], chemical remote sensing [16] and constrained energy minimization (CEM) in hyperspectral image classification [17–19]. Of particular interest is a comparative study between LCDA and CEM since CEM has shown success in some practical applications. The advantage of CEM over LCDA is that CEM is designed for detection of a particular target and only requires knowledge of the desired target to be detected and classified. So, it is very sensitive to noise. On the other hand, LCDA needs a complete knowledge of target signatures of interest, but can do much better classification than CEM using direction constraints when two targets have very similar signatures in which CEM can generally detect one of them, but not both. Furthermore, the advantage of using CEM in unknown image scenes diminishes when LCDA and CEM are extended to their unsupervised versions where the exact target knowledge is not available.

This paper is organized as follows. Section 2 briefly reviews Fisher's LDA and CEM. Section 3 describes LCDA in detail. In particular, an algorithm to implement LCDA is also provided in this section. Section 4 extends LCDA, LDA, CEM and OSP to their unsupervised versions. Section 5 conducts experiments using airborne visible/infrared imaging spectrometer (AVIRIS) and HYperspectral Digital Imagery Collection Experiment (HYDICE) data to evaluate the performance of LCDA and ULCDA in comparison with LDA, CEM and OSP. Finally, Section 6 draws some conclusions.

2. Fisher's linear discriminant analysis and constrained energy minimization

In this section, we briefly review Fisher's linear discriminant analysis (LDA) and constrained energy minimization (CEM) approaches which will be used in comparison with LCDA.

2.1. Fisher's linear discriminant analysis (LDA)

Let R^d and R^p be d- and p-dimensional vector spaces, respectively, with $p \leqslant d$. Let $\{C_k\}_{k=1}^{c}$ denote c classes of interest where $C_k = \{\mathbf{x}_1^k, \mathbf{x}_2^k, \ldots, \mathbf{x}_{N_k}^k\} = \{\mathbf{x}_j^k\}_{j=1}^{N_k}$ is the kth class and contains N_k patterns and the jth pattern in class C_k, denoted by $\mathbf{x}_j^k = (x_{1j}^k\, x_{2j}^k \ldots x_{dj}^k)^{\mathrm{t}}$ is a d-dimensional vector in the space R^d. Let $N = N_1 + \cdots + N_c$ be the total number of training patterns. From Fisher's discriminant analysis [10], we can form total, between- and within-class scatter matrices as follows.

Let $\mu = (1/N)\sum_{k=1}^{N}\sum_{j=1}^{N_k}\mathbf{x}_j^k$ be the global mean and $\mu_k = (1/N_k)\sum_{j=1}^{N_k}\mathbf{x}_j^k$ the mean of class C_k. Then

$$S_{\mathrm{T}} = \sum_{k=1}^{c}\sum_{j=1}^{N_k}(\mathbf{x}_j^k - \mu)(\mathbf{x}_j^k - \mu)^{\mathrm{t}}, \tag{1}$$

$$S_{\mathrm{W}} = \sum_{k=1}^{c}\sum_{j=1}^{N_k}(\mathbf{x}_j^k - \mu_k)(\mathbf{x}_j^k - \mu_k)^{\mathrm{t}}, \tag{2}$$

$$S_{\mathrm{B}} = \sum_{k=1}^{c} N_k(\mu_k - \mu)(\mu_k - \mu)^{\mathrm{t}}. \tag{3}$$

From Eqs. (1)–(3)

$$S_{\mathrm{T}} = S_{\mathrm{W}} + S_{\mathrm{B}}. \tag{4}$$

Assume that $\Xi = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} = \{\mathbf{x}_j^k\}_{j=1,k=1}^{N_k,c}$ are all data training samples. The Fisher's discriminant analysis is to find a $d \times (c-1)$ weight matrix $W = [\mathbf{w}_1\ \mathbf{w}_2 \cdots \mathbf{w}_{c-1}]$ that projects all data samples $\mathbf{x} \in \Xi$ in an R^d space into

$\mathbf{y}$ in a low-dimensional feature space R^p in such a manner that all projected data samples $\mathbf{y}$'s yield the best-possible class separability by

$$y = W^{\mathrm{t}}\mathbf{x} \tag{5}$$

with

$$y_k = \mathbf{w}_k^{\mathrm{t}}\mathbf{x} \quad \text{for} \quad 1 \leqslant k \leqslant c-1, \tag{6}$$

where $\mathbf{w}_k$ is the kth column vector with dimensionality $d \times 1$ in W and $\mathbf{y} = (\mathbf{y}_1 \mathbf{y}_2 \cdots \mathbf{y}_N)^{\mathrm{t}}$.

Using Eqs. (2) and (3) we can define similar within- and between-class scatter matrices for the projected samples $\mathbf{y}$ given by (5) as follows:

$$\tilde{S}_{\mathrm{W}} = \sum_{k=1}^{c} \sum_{j=1}^{N_k} (\mathbf{y}_j^k - \tilde{\mu}_k)(\mathbf{y}_j^k - \tilde{\mu}_k)^{\mathrm{t}}, \tag{7}$$

$$\tilde{S}_{\mathrm{B}} = \sum_{k=1}^{c} N_k (\tilde{\mu}_k - \tilde{\mu})(\tilde{\mu}_k - \tilde{\mu})^{\mathrm{t}}. \tag{8}$$

where $\tilde{\mu} = (1/N)\sum_{k=1}^{N}\sum_{j=1}^{N_k} y_j^k$ and $\tilde{\mu}_k = (1/N_k)\sum_{j=1}^{N_k} y_j^k$.

Substituting Eqs. (2), (3), (5) and (6) into Eqs. (7) and (8) results in

$$\tilde{S}_{\mathrm{W}} = W^{\mathrm{t}} S_{\mathrm{W}} W, \tag{9}$$

$$\tilde{S}_{\mathrm{B}} = W^{\mathrm{t}} S_{\mathrm{B}} W, \tag{10}$$

In order to find an optimal linear transformation matrix W in the sense of class separability, we use Fisher's discriminant function ratio, called Raleigh quotient which is the ratio of the between-class scatter to within-class scatter as follows:

$$J(W) = \frac{|\tilde{S}_{\mathrm{B}}|}{|\tilde{S}_{\mathrm{W}}|} = \frac{|W^{\mathrm{t}} S_{\mathrm{B}} W|}{|W^{\mathrm{t}} S_{\mathrm{W}} W|}, \tag{11}$$

where $|\cdot|$ is the determinant of a matrix. The optimal solution to Eq. (11), denoted by

$$W^*_{d\times(c-1)} = [\mathbf{w}_1^* \mathbf{w}_2^* \cdots \mathbf{w}_{c-1}^*]_{d\times(c-1)} \tag{12}$$

can be found by solving following generalized eigenvalue problem:

$$S_{\mathrm{B}}\mathbf{w}_k^* = \lambda_k S_{\mathrm{W}} \mathbf{w}_k^* \tag{13}$$

with $\mathbf{w}_k^*$ corresponding to the eigenvalue λ_k. These $c-1$ eigenvectors, $\{\mathbf{w}_k^*\}_{k=1}^{c-1}$ form a set of Fisher's linear discriminant functions which can be used in Eq. (6) as

$$y_k = (\mathbf{w}_k^*)^{\mathrm{t}}\mathbf{x} \quad \text{for } 1 \leqslant k \leqslant c-1. \tag{14}$$

It should be noted that since there are c classes, only $c-1$ eigenvalues, denoted by $\{\lambda_i\}_{i=1}^{c-1}$ are nonzeros. Each eigenvalue λ_i generates its own eigenvector $\mathbf{w}_j^*$. By means of these eigenvector $\{\mathbf{w}_i^*\}_{i=1}^{c-1}$ we can define a Fisher's discriminant analysis-based optimal linear transformation T^* via Eqs. (5), (12) and (14) by

$$y = T^*(\mathbf{x}) = (W^*_{d\times(c-1)})^{\mathrm{t}}\mathbf{x}. \tag{15}$$

For more details, we refer to Ref. [10].

2.2. *Constrained energy minimization* (*CEM*) [17]

An approach similar to direction constraints, called constrained energy minimization (CEM) [17–19] was previously developed for detection and classification of a desired target. It used a finite impulse response (FIR) filter to constrain the desired signature by a specific gain while minimizing the filter output power. The idea of CEM arises in minimum variance distortionless response (MVDR) in array processing [14,15] with the desired signature interpreted as the signals arrived from a desired direction. It can be derived as follows.

Assume that we are given a finite set of observations $S = \{\mathbf{r}_1 \mathbf{r}_2 \cdots \mathbf{r}_N\}$ where $\mathbf{r}_i = (r_{i1} r_{i2} \cdots r_{iL})^{\mathrm{t}}$ for $1 \leqslant i \leqslant N$ is a sample pixel vector. Suppose that the desired signature $\mathbf{d}$ is also known a priori. The objective of CEM is to design an FIR linear filter with L filter coefficients $\{w_1 w_2 \cdots w_L\}$, denoted by an L-dimensional vector $\mathbf{w} = (w_1 w_2 \cdots w_L)^{\mathrm{t}}$ that minimizes the filter output power subject to the following constraint:

$$\mathbf{d}^{\mathrm{t}}\mathbf{w} = \sum_{l=1}^{L} d_l w_l = 1. \tag{16}$$

Let y_i denote the output of the designed FIR filter resulting from the input $\mathbf{r}_i$. Then y_i can be written as

$$y_i = \sum_{l=1}^{L} w_l r_{il} = \mathbf{w}^{\mathrm{t}}\mathbf{r}_i = \mathbf{r}_i^{\mathrm{t}}\mathbf{w}. \tag{17}$$

So, the average output power produced by the observation set S and the FIR filter with coefficient vector $\mathbf{w} = (w_1 w_2 \cdots w_L)^{\mathrm{t}}$ specified by Eq. (17) is given by

$$\frac{1}{N}\left[\sum_{i=1}^{N} y_i^2\right] = \frac{1}{N}\left[\sum_{l=1}^{L} (\mathbf{r}_i^{\mathrm{t}}\mathbf{w})^{\mathrm{t}}\mathbf{r}_i^{\mathrm{t}}\mathbf{w}\right] = \mathbf{w}^{\mathrm{t}}\left(\frac{1}{N}\left[\sum_{i=1}^{N}\mathbf{r}_i\mathbf{r}_i^{\mathrm{t}}\right]\right)\mathbf{w} = \mathbf{w}^{\mathrm{t}}R_{L\times L}\mathbf{w}, \tag{18}$$

where $R_{L\times L} = (1/N)\left[\sum_{i=1}^{N}\mathbf{r}_i\mathbf{r}_i^{\mathrm{t}}\right]$ turns out to be the $L \times L$ sample autocorrelation matrix of S.

Minimizing Eq. (18) with the filter response constraint $\mathbf{d}^{\mathrm{t}}\mathbf{w} = \sum_{l=1}^{L} d_l w_l = 1$ yields

$$\min_{w}\left\{\frac{1}{N}\left[\sum_{i=1}^{N} y_i^2\right]\right\} = \min_{w}\{\mathbf{w}^{\mathrm{t}}R_{L\times L}\mathbf{w}\} \text{ subject to } \mathbf{d}^{\mathrm{t}}\mathbf{w} = 1. \tag{19}$$

The solution to Eq. (19) was shown [17] and called constrained energy minimization (CEM) classifier with the weight vector $\mathbf{w}^*$ given by

$$\mathbf{w}^* = \frac{R_{L\times L}^{-1}\mathbf{d}}{\mathbf{d}^t R_{L\times L}^{-1}\mathbf{d}}. \tag{20}$$

3. Linear constrained distance-based discriminant analysis

In Fisher's discriminant analysis, the discriminant functions are generated based on the Fisher's ratio with no constraints on directions of these discriminant functions. In many practical applications, a priori knowledge can be used to constrain desired features along certain directions to minimize the effects of undesired features. As discussed in Section 2.2 CEM constrains a desired signature with a specific gain so that its output energy is minimized. In MRI [11], Soltanian-Zadeh et al. constrained normal tissues along prespecified target positions so as to achieve a better clustering for visualization. In this section, we follows a similar idea [11] to derive a constrained discriminant analysis for hyperspectral image classification.

First of all, assume that a linear transformation T is used to project a high-dimensional data space into a low-dimensional space using the ratio of the average between-class distance to average within-class distance as the criterion for optimality [11] and given by

$$J(T) = \frac{(2/c(c-1))\sum_{i=1}^{c}\sum_{j=i+1}^{c}\|T(\mu_i)-T(\mu_j)\|^2}{(1/cN)\sum_{k=1}^{c}[\sum_{j=1}^{N_k}\|T(\mathbf{x}_j^k)-T(\mu_k)\|^2]}, \tag{21}$$

where the global mean μ and local mean μ_k of the kth class were defined in the previous section. Now suppose that there are p classes of particular interest with $p \leqslant c$ which are denoted by $\{C_k\}_{k=1}^{p}$ without loss of generality. We can implement Eq. (21) subject to constraint that the desired class means $\{\mu_k\}_{k=1}^{p}$ along the prespecified target directions $\{\mathbf{t}_k\}_{k=1}^{p}$. What we seek here is to find an optimal linear transformation T^* which maximizes

$$J(T) \text{ subject to } \mathbf{t}_k = T(\mu_k) \text{ for all } k \text{ with } 1 \leqslant k \leqslant p \leqslant c. \tag{22}$$

Eqs. (21) and (22) outline a general constrained optimization problem which can be solved numerically. Like Eq. (15) we can view the linear transformation T in Eq. (22) as a matrix transform which is specified by a weight matrix $W_{d\times p}$

$$\mathbf{y} = T(\mathbf{x}) = W_{d\times p}^{t}\mathbf{x}, \tag{23}$$

where $W_{d\times p} = [\mathbf{w}_1\mathbf{w}_2\cdots\mathbf{w}_p]_{d\times p}$ is given similarly by Eq. (12).

Now if we further assume that $p = c$, $\mathbf{t}_k$, in Eq. (22) is a unit vector along the kth coordinate in the space R^p, i.e., $\mathbf{t}_k = (0\cdots 0\underset{k}{1}\,0\cdots 0)^{\mathrm{T}}$ is a $p\times 1$ unit column vector with one in the kth component and zeros, otherwise and $\{\mathbf{w}_i\}_{i=1}^{p}$ in $W_{d\times p}$ are linearly independent, then Eq. (22) becomes

$$J(T) \text{ subject to } \mathbf{w}_i^t\mu_k = \delta_{ik} \text{ for } 1 \leqslant j,k \leqslant p, \tag{24}$$

where δ_{jk} is Kronecker's notation given by

$$\delta_{ik} = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{if } i \neq k. \end{cases}$$

From Ref. [20], the numerator and denominator of $J(T)$ in Eq. (21) can be further reduced to

$$\frac{2}{p(p-1)}\sum_{i=1}^{p}\sum_{j=i+1}^{p}\|T(\mu_i)-T(\mu_j)\|^2 = \frac{2}{p(p-1)}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}[2] = 2 \tag{25}$$

and

$$\frac{1}{pN}\sum_{k=1}^{p}\left[\sum_{j=1}^{N_k}\|T(\mathbf{x}_j^k)-T(\mu_k)\|^2\right] = \frac{1}{pN}\,trace\left(W^t\left\{\sum_{k=1}^{p}\left[\sum_{j=1}^{N_k}(\mathbf{x}_j^k-\mu_k)(\mathbf{x}_j^k-\mu_k)^t\right]\right\}W\right). \tag{26}$$

Since the numerator given by Eq. (25) is a constant, maximizing Eq. (24) is reduced to minimizing Eq. (26), namely

$$\min_{w}\ trace\left(W^t\left\{\sum_{k=1}^{p}\left[\sum_{j=1}^{N_k}(\mathbf{x}_j^k-\mu_k)(\mathbf{x}_j^k-\mu_k)^t\right]\right\}W\right) \quad \text{subject to } \mathbf{w}_i^t\mu_k = \delta_{ik}. \tag{27}$$

It should be noted that the term in the bracket in Eq. (27) turns out to be S_{W}. From Eq. (4) minimizing S_{W} is equivalent to minimizing S_{T}. So, if let Σ denote the sample covariance matrix of all training samples, then

$$S_{\mathrm{T}} = N\cdot\Sigma. \tag{28}$$

Since N is a constant, using Eq. (28) we obtain the following equivalent problem by substituting Eq. (28) into Eq. (27):

$$\min_{W}\ trace(W^t\Sigma W) \text{ subject to } \mathbf{w}_i^t\mu_k = \delta_{ik}. \tag{29}$$

Now all problems are reduced to how we can find a matrix A which decorrelates and whitens the covariance matrix Σ into an identity matrix so that the Gram–Schmidt orthogonalization procedure can be employed to further orthogonalize all $A^t\mu_k$'s. Assume that there exists such a matrix A. Eq. (29) can be further reduced to a very simple optimization problem given by

$$\min_W \; trace(W^t \Sigma W) = \min_W \left\{ \sum_{i=1}^{p} \mathbf{w}_i^t \mathbf{w}_i \right\}$$
$$= \min_W \left\{ \sum_{i=1}^{p} \|\mathbf{w}_i\|^2 \right\} \text{subject to } \mathbf{w}_i^t \hat{\mu}_k = \delta_{ik}, \quad (30)$$

where $\{\hat{\mu}_i\}_{i=1}^p$ is the orthogonal vectors resulting from applying Gram–Schmidt orthogonalization procedure to $A^t \mu_k$. Since $\|\mathbf{w}_i\|^2$ is nonnegative for each i with $1 \leqslant i \leqslant p$, Eq. (30) is also equivalent to

$$\max_{w_i} (\mathbf{w}_i^t \mathbf{w}_i)^{-1/2} \text{ subject to } \mathbf{w}_i^t \hat{\mu}_k = \delta_{ik}$$
$$\text{for each } i \text{ with } 1 \leqslant i \leqslant p \quad (31)$$

which can be solved analytically [11] for each $\mathbf{w}_i^*$ given by

$$\mathbf{w}_i^* = \hat{\mu}_i^t P_U^{\perp}, \quad (32)$$

where

$$P_U^{\perp} = I - U(U^t U)^{-1} U^t \quad (33)$$

and U is the space linearly spanned by $\{\hat{\mu}_j\}_{j=1, j \neq i}^p$, all cluster centers except $\hat{\mu}_i$. Surprisingly, the solution specified by Eq. (32) turns out to be the orthogonal subspace projection classifier [7]. So, the classifier $\mathbf{w}_i^* = \hat{\mu}_i^t P_U^{\perp}$ in LCDA can be viewed as a constrained version of the OSP classifier.

A comment on LCDA is noteworthy. Despite the fact that two criteria used in LDA and LCDA look similar where both calculate a ratio of a measure of between class to a measure of within class, there are differences between LDA and LCDA. In LDA, it uses the scatter matrices derived from between class and within class while LCDA makes use of inter-distance between classes and intra-distance within classes to arrive the criterion specified by Eq. (21). Another difference also noted in [11] is that the number of the discriminant functions resulting from LDA is one less than the total number of classes of interest, p, whereas the number of projection vectors used in LCDA is the same number of classes of interest, p. Most significantly, LCDA can be viewed as a variation of OSP as shown above but LDA is not.

4. Implementation of LCDA using a whitening process

As mentioned, in order to reduce the optimization problem given by Eq. (29) to the one specified by Eq. (30), we need to find the matrix A. In Ref. [11], Soltanian-Zadeh et al. made the white Gaussian noise assumption for MRI to arrive at Eq. (30). As a matter of fact, we can find the matrix A without making such an assumption.

Assume that $\{\lambda_i\}_{i=1}^d$ are the eigenvalues of the sample covariance matrix Σ or the total scatter matrix S_T and $\{\mathbf{v}_i\}_{i=1}^d$ are their corresponding eigenvectors. Since S_T or Σ is nonnegative definite, then all eigenvalues are nonnegative and there exists a unitary matrix Q such that Σ can be decomposed into

$$Q^t \Sigma Q = \Lambda, \quad (34)$$

where $Q = [\mathbf{v}_1 \mathbf{v}_2 \cdots \mathbf{v}_b]$ is a matrix made up of the eigenvectors $\{\mathbf{v}_i\}_{i=1}^d$ and $\Lambda = diag\{\lambda_i\}_{i=1}^d$ is a diagonal matrix with $\{\lambda_i\}_{i=1}^d$ in the diagonal line.

If we let $\Lambda^{-1/2} = diag\{\sqrt{\lambda_i}\}_{i=1}^d$, multiplying both sides of Eq. (34) by $\Lambda^{-1/2}$ results in

$$\Lambda^{-1/2} Q^t \Sigma Q \Lambda^{-1/2} = I. \quad (35)$$

From Eq. (35), we obtain the desired matrix A for Eq. (29) which is given by

$$A = Q\Lambda^{-1/2} \quad (36)$$

so that $A^t \Sigma A = I$.

Using Eqs. (36) and (32) we can solve the linear constrained Euclidean distance-based discriminant analysis optimization problem specified by Eq. (30) as follows:

Algorithm to Implement LCDA

1. Find the cluster centers $\{\mu_k\}_{k=1}^p$ of p-classes and the total scatter matrix S_T or covariance matrix Σ.
2. Find the eigenvalues $\{\lambda_i\}_{i=1}^d$ and their corresponding eigenvectors $\{\mathbf{v}_i\}_{i=1}^d$ of S_T or Σ to form the unitary matrix $Q = [\mathbf{v}_1 \mathbf{v}_2 \cdots \mathbf{v}_d]$ and the diagonal matrix $\Lambda = diag\{\lambda_i\}_{i=1}^d$.
3. Form the desired matrix $A = Q\Lambda^{-1/2}$ using Eq. (36).
4. Find the A-transformed cluster centers $\{A^t \mu_i\}_{i=1}^p$ where for each i with $1 \leqslant i \leqslant p$.
5. Apply the Gram–Schmidt orthogonalization procedure to orthogonalize $\{A^t \mu_i\}_{i=1}^p$ to produce their corresponding orthogonal vectors $\{\hat{\mu}_i\}_{i=1}^p$.
6. The linear constrained discriminant Euclidean distance-based analysis optimization problem specified by Eq. (30) can be solved by $\mathbf{w}_i^* = \hat{\mu}_i^t P_U^{\perp}$ with $P_U^{\perp}$ given by Eq. (33). This step is the classification step where $\mathbf{w}_i^*$ is used to classify data samples into the ith class.

5. Unsupervised LCDA

LCDA described in Section 3 requires training samples to generate data sample covariance matrix Σ. In order to extend LCDA to unsupervised LCDA many unsupervised clustering methods can be used for this purpose, for instance, ISODATA, K-means clustering [10,12]. In hyperspectral image classification, we can design an unsupervised LCDA by taking advantage of an algorithm, called target generation process (TGP) used in the automatic target detection and classification algorithm [21–23].

The idea of TGP can be briefly described as follows. It is first initialized by selecting a pixel vector with the maximum length as an initial target denoted by $\mathbf{T}_0$. We then employ an orthogonal subspace projector $P^{\perp}_{\mathbf{T}_0}$ via Eq. (33) with $U = \mathbf{T}_0$ to project all image pixel vectors into the orthogonal complement space of $\langle \mathbf{T}_0 \rangle$, denoted by $\langle \mathbf{T}_0 \rangle^{\perp}$. The maximum length of the pixel vector in $\langle \mathbf{T}_0 \rangle^{\perp}$ will be selected to be a first target denoted by $\mathbf{T}_1$. The reason for this selection is the following. Since $\mathbf{T}_1$ has the maximum projection in $\langle \mathbf{T}_0 \rangle^{\perp}$, $\mathbf{T}_1$ should have the most distinct features from $\mathbf{T}_0$. Then we create the first set U_1 by letting $U_1 = \mathbf{T}_1$ and calculate the Orthogonal Projection Correlation Index (OPCI) defined by $\eta_1 = T^t_0 P^{\perp}_{(\mathbf{T}_0 \cdot U_1)} \mathbf{T}_0$ to see whether it less than a prescribed threshold. If it is, TGP stops generating targets. Otherwise, TGP is continued to search for a second target by applying $P^{\perp}_{(\mathbf{T}_0 \mathbf{T}_1)}$ again to the image. The pixel vector with the maximum length in the space $\langle \mathbf{T}_0, \mathbf{T}_1 \rangle^{\perp}$ will be selected as a second target, denoted by $\mathbf{T}_2$. Then once again, we calculate the $\eta_2 = \mathrm{OPCI}(\mathbf{T}_0, U_2) = \mathbf{T}^t_0 P^{\perp}_{(\mathbf{T}_0 \cdot U_2)} \mathbf{T}_0$ with $U_2 = (\mathbf{T}_1 \mathbf{T}_2)$ to determine if TGP should be terminated. If not, the same above procedure is repeated to find a third target, a fourth target, etc. until at the ith step where the $\eta_i = \mathrm{OPCI}(\mathbf{T}_0, U_i) = \mathbf{T}^t_0 P^{\perp}_{(\mathbf{T}_0 \cdot U_i)} \mathbf{T}_0$ is small enough and less than a prescribed threshold.

The objective of TGP is to generate a set of potential targets in an unknown image without any prior knowledge. Because of that each TGP-generated target may not be a real target and could be its neighboring pixel vectors due to interfering effects. In order to produce a robust signature matrix M, several metrics can be used for this purpose such as Euclidean distance (ED) [10], spectral angle (SA) [1], spectral information divergence (SID) [24,25]. Assume that $\{\mathbf{T}_k\}^c_{k=1}$ is the set of targets generated by TGP and $m(x, y)$ is a metric to measure the closeness of two samples x and y. Then for each $1 \leqslant k \leqslant c$ the training class of $\mathbf{T}_k$, denoted by C_i comprises all pixel vectors with distance from $\mathbf{T}_j$ measured by $m(\cdot,\cdot)$ less than a prescribed threshold ε. That is

$$C_k = \{\mathbf{x}^k | m(\mathbf{x}^k, \mathbf{T}_k) < \varepsilon\}, \tag{37}$$

where ε is a prescribed distance threshold. It should be noted that $\mathbf{T}_k$ is included in its own class C_k. By virtue of these training classes $\{C_k\}^c_{k=1}$, ULCDA can be implemented by LCDA.

6. Experimental results

In this section, hyperspectral data will be used to evaluate the performance of LCDA and ULCDA in comparison with the results produced by OSP [7] and CEM [17–19].

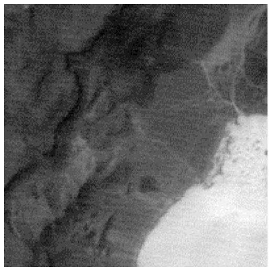


Fig. 1. An AVIRIS image scene.

Example 1 (*AVIRIS experiments*). The AVIRIS data used in the experiments were the same data considered in Ref. [7]. It is a subscene of 200×200 pixels extracted from the upper left corner of the Lunar Crater Volcanic Field in Northern Nye County, Nevada shown in Fig. 1 where five signatures of interest in these images are “red oxidized basaltic cinders”, “rhyolite”, “playa (dry lakebed)”, “shade” and “vegetation”. In this case, $p = 5$ and $d = 224$ is the number of bands. LCDA used constrained unit vectors $\{\mathbf{t}_k\}^5_{k=1}$ in Eq. (22) to steer the five desired signatures along five orthogonal directions. Fig. 2 shows results of LCDA, CEM and OSP where the images in the first, second, third and fourth columns were produced by LCDA, LDA, CEM and OSP, respectively. Images labeled by (a), (b), (c) and (d) show targets: cinders, rhyolite, playa and vegetation as targets, respectively, and images labeled by (e) are results of the shade. The images are arranged in such a fashion that their counterparts can be compared in parallel. As we can see from Fig. 2, the results produced by LCDA and CEM are comparable and both performed better than LDA and OSP in detection and classification of all five target signatures, cinders, rhyolite, playa and vegetation, specifically, in classifying cinders, vegetation and shade. Here, the LDA was carried out by performing LDA on the image data then followed by a minimum-distance classifier used for class-membership assignment. Since it worked as a pure pixel classifier rather than a mixed pixel classifier to estimate target signature abundance as performed by the other three the images produced by LDA are binary as opposed to gray scale images produced by LCDA, CEM and OSP. Therefore, it is not surprising to see that LDA produced the worst results.

In order to see the performance of ULCDA, we adopted SID [24,25] as a spectral metric to measure the closeness or similarity of two pixel vectors in the scene. The unsupervised clustering used in ULCDA was the nearest neighboring rule (NNR) [10,12]. We also

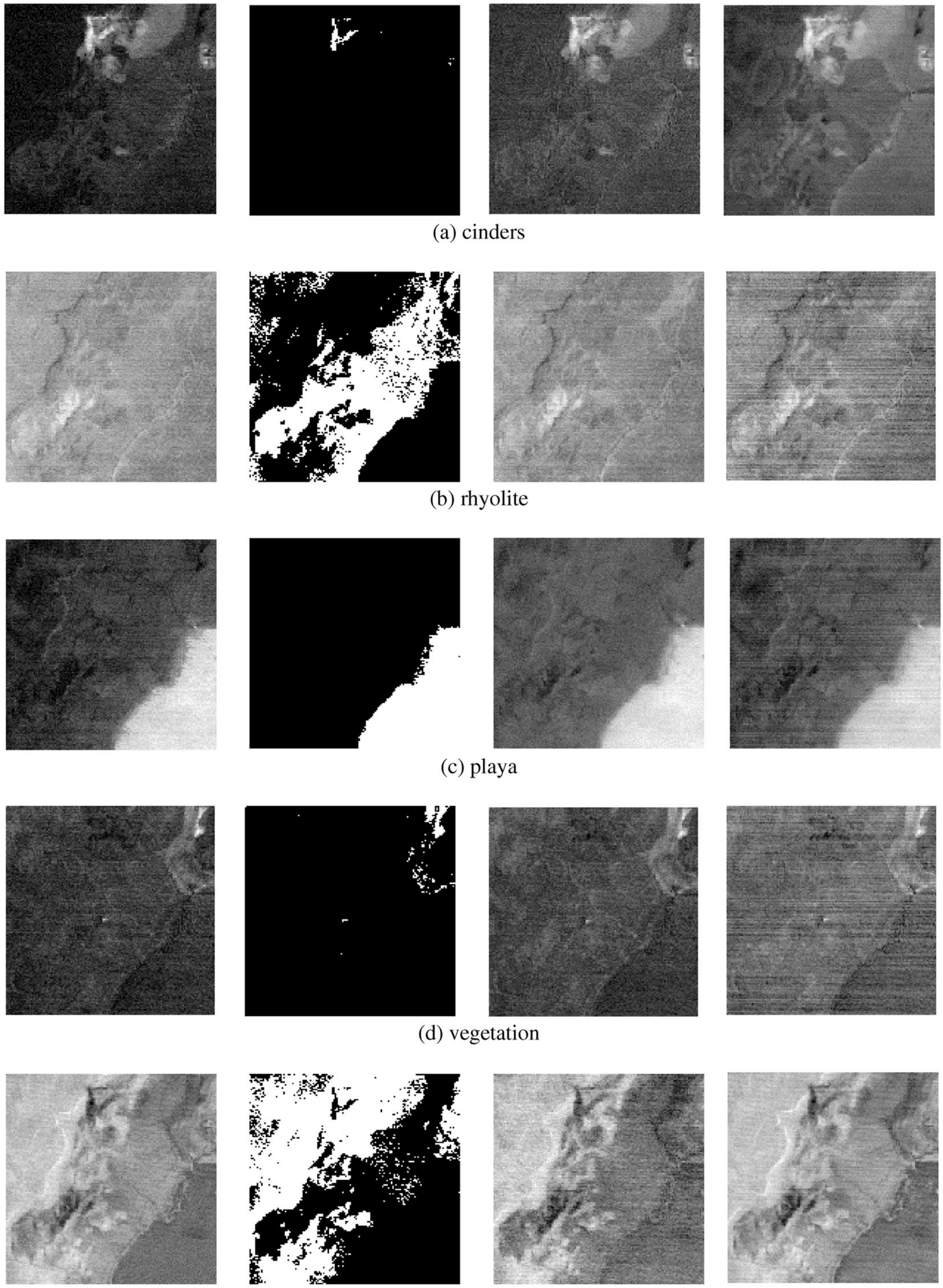


Fig. 2. Results of LCDA, LDA CEM and OSP. First column: (a) cinders classified by LCDA; (b) rhyolite classified by LCDA; (c) playa classified by LCDA; (d) vegetation classified by LCDA; (e) shade classified by LCDA. Second column: (a) cinders classified by LDA; (b) rhyolite classified by LDA; (c) playa classified by LDA; (d): vegetation classified by LDA; (e) shade classified by LDA. Third column (a) cinders classified by CEM; (b); rhyolite classified by CEM; (c) playa classified by CEM; (d) vegetation classified by CEM; (e) shade classified by CEM. Forth column: (a) cinders classified by OSP; (b) rhyolite classified by OSP; (c): playa classified by OSP; (d): vegetation classified by OSP; (e): shade classified by OSP.

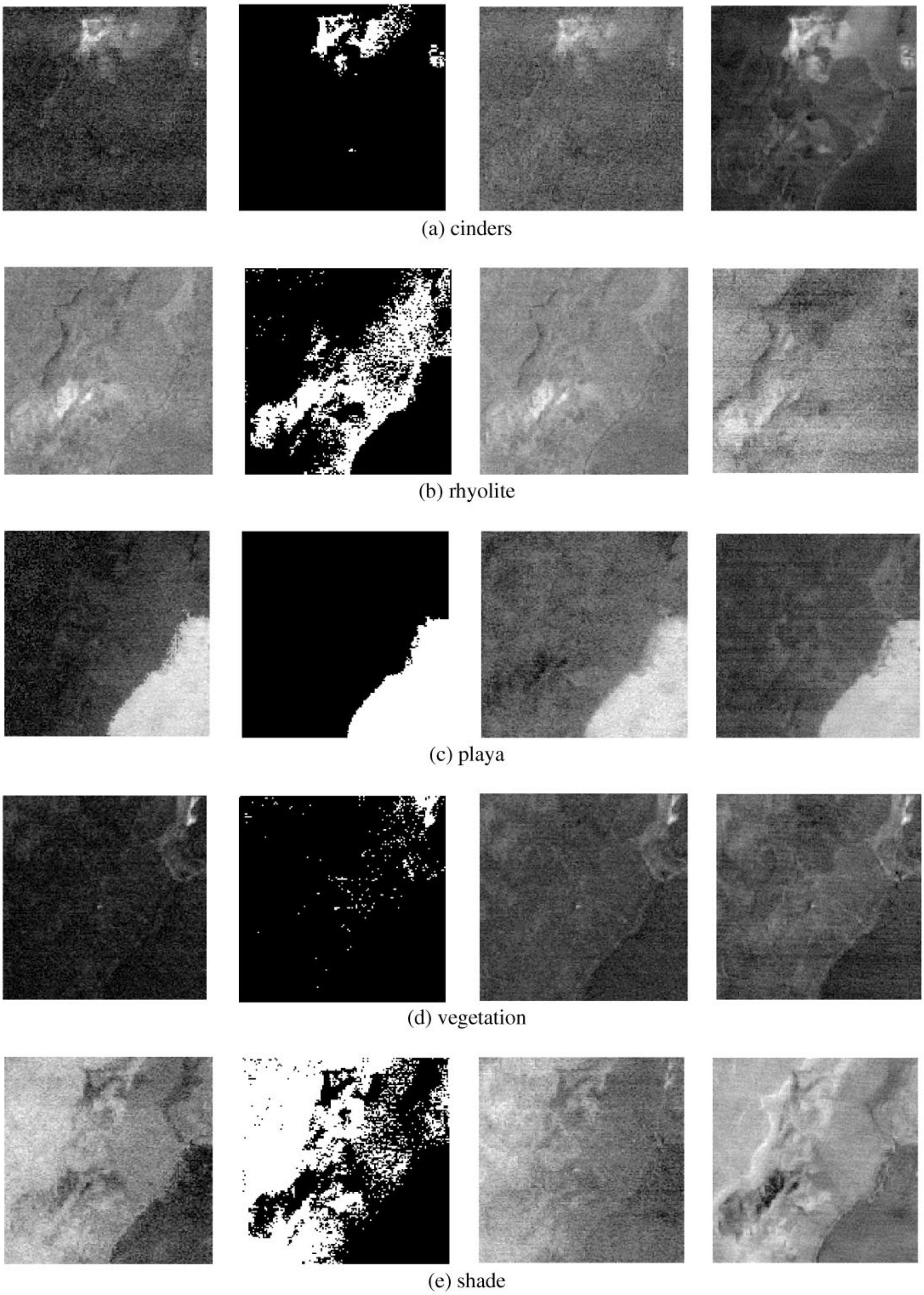


Fig. 3. Results of ULCDA, CEM and UOSP with SID. First column: (a) cinders classified by ULCDA; (b) rhyolite classified by ULCDA; (c) playa classified by ULCDA; (d) vegetation classified by ULCDA; (e) shade classified by ULCDA. Second column: (a) cinders classified by ULDA; (b) rhyolite classified by ULDA; (c) playa classified by ULDA; (d): vegetation classified by ULDA; (e) shade classified by ULDA. Third column: (a) cinders classified by UCEM; (b): rhyolite classified by UCEM; (c): playa classified by UCEM; (d): vegetation classified by UCEM; (e): shade classified by UCEM. Forth column: (a) cinders classified by UOSP; (b): rhyolite classified by UOSP; (c) playa classified by UOSP; (d) vegetation classified by UOSP; (e) shade classified by UOSP.

used TGP, SID and NNR to extend LDA, OSP and CEM to their unsupervised counterparts, referred to as ULDA, UCEM and UOSP. The purpose of using SID is to relax the sensitivity of knowledge used in classification. Since there is no prior knowledge about signatures, the target pixel vectors generated by TGP may not be true target vectors due to interference and noise. Using SID in Eq. (37) allows us to group pixel vectors whose signatures are similar and close to desired target vectors and use their spectral signature averages as target class centers. The results are shown in Fig. 3. All the images are arranged in the same matter as does Fig. 2. It is interesting to note that in Fig. 3, except for ULDA that did much worse than LDA, the classification results obtained by ULCDA, UCEM and UOSP are nearly the same as those obtained by their counterparts in Fig. 2 even there is no a priori signature knowledge is available. The reason for this is that the a priori target signature information required for mixed pixel classification can be compensated by its estimated abundance while ULDA does not have such an advantage.

Example 2 (*HYDICE experiments*). In this example, a HYDICE scene was conducted to evaluate the performance among LCDA, LDA, CEM and OSP and their unsupervised counterparts. The considered scene is exactly the same one used in [13] and reproduced in Fig. 4 with 1.5-m spatial resolution. Four vehicles of two different types are parked along the tree line labeled from top to bottom by V_1, V_2, V_3, V_4 and one man-made object denoted by Obj is near the center of the scene. While the top three V_1, V_2, V_3 belong to one type of vehicle denoted by V1, the bottom vehicle V_4 belongs to another type of vehicle denoted by V2. Of particular interest in this scene is that the ground truth provides precise pixel locations of all the four vehicles as well as the object so that we can verify the detection and classification results for each target for different methods. Fig. 5 shows the results of LCDA, LDA, CEM and OSP with the images arranged in the same fashion as those in Fig. 2. Images labeled by (a), (b) and (c) show the target detection and classification of the Obj, V1 and V2, respectively. From Fig. 5 LCDA performed better than LDA, CEM and OSP in overall performance. An interesting finding is that LDA actually performed better than LCDA in detection of V_4 but did not work as well as LCDA in detection of V1 and Obj where many false alarms occurred in LDA detection. The reason for this is that LDA is a pure pixel classification technique while LCDA is a mixed pixel classification method (So are CEM and OSP). As a result, the gray level values of mixed pixels in the images generated by LCDA reflect the abundance fractions of a particular detected target. Due to the fact that the spectral signature of V_3 is very similar to that of V_4, LCDA also detected a very small fraction of V_3 while detecting V_4 in Fig. 5(c).

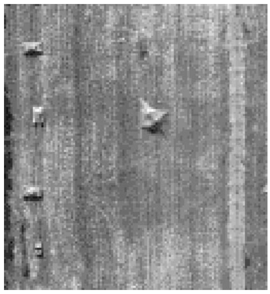


Fig. 4. A HYDICE scene.

The same phenomena were found much worse in the CEM-generated and OSP-generated images shown in Fig. 5(b) and (c) where both methods had difficulty with differentiating these two vehicles V_3 and V_4. Unlike CEM and OSP, LCDA did manage to mitigate this problem by making use of constraints to steer V_3 and V_4 in such a way that they both were forced to be separated along orthogonal directions. Consequently, a barely visible amount of abundance of V_3 was detected and classified in the LCDA-detected image in Fig. 5(c). Furthermore, comparing the images in the first and third columns of Fig. 5, we can see that CEM extracted more abundance of V_3 than did LCDA in detection of V_4. Nevertheless, CEM and LCDA performed significantly better than OSP. In detection and classification of V1, LCDA also performed better than LDA, CEM and OSP as shown in the images of Fig. 5. Although OSP also correctly classified V1, it also extracted some natural background signatures, tree and road. On the contrary, CEM nulled out all background signatures, but also inevitably extracted some fraction of V_4. Since the spectral signature of Obj is very distinct from those of four vehicles, OSP, CEM and LCDA performed well in this case. These HYDICE experiments further demonstrate that in discrimination of targets with similar spectral signatures LCDA not only is superior to OSP, but also performs slightly better than CEM.

As noted above, the spectral signatures of V_3 and V_4 are very similar. So, when ULCDA, ULDA, UCEM and UOSP were applied to the scene in Fig. 4, the results were interesting. Since there is no a priori knowledge about the targets, V_3 and V_4 were treated as different targets. As a results, four categories of targets, Obj, V1(V_1, V_2), V1(V_3), V2(V_4) were detected in Fig. 6 where the images are also arranged in the same way as those in Fig. 5 but V1 has been split into two different classes. As shown in Fig. 6, ULCDA did not work as well as did LCDA in Fig. 5, but still performed reasonably well in general. ULDA did well in pulling out Obj and V1(V_3), but did poorly in detecting V1(V_1, V_2) and V2(V_4) with many false alarms. Of particular interest is UCEM where

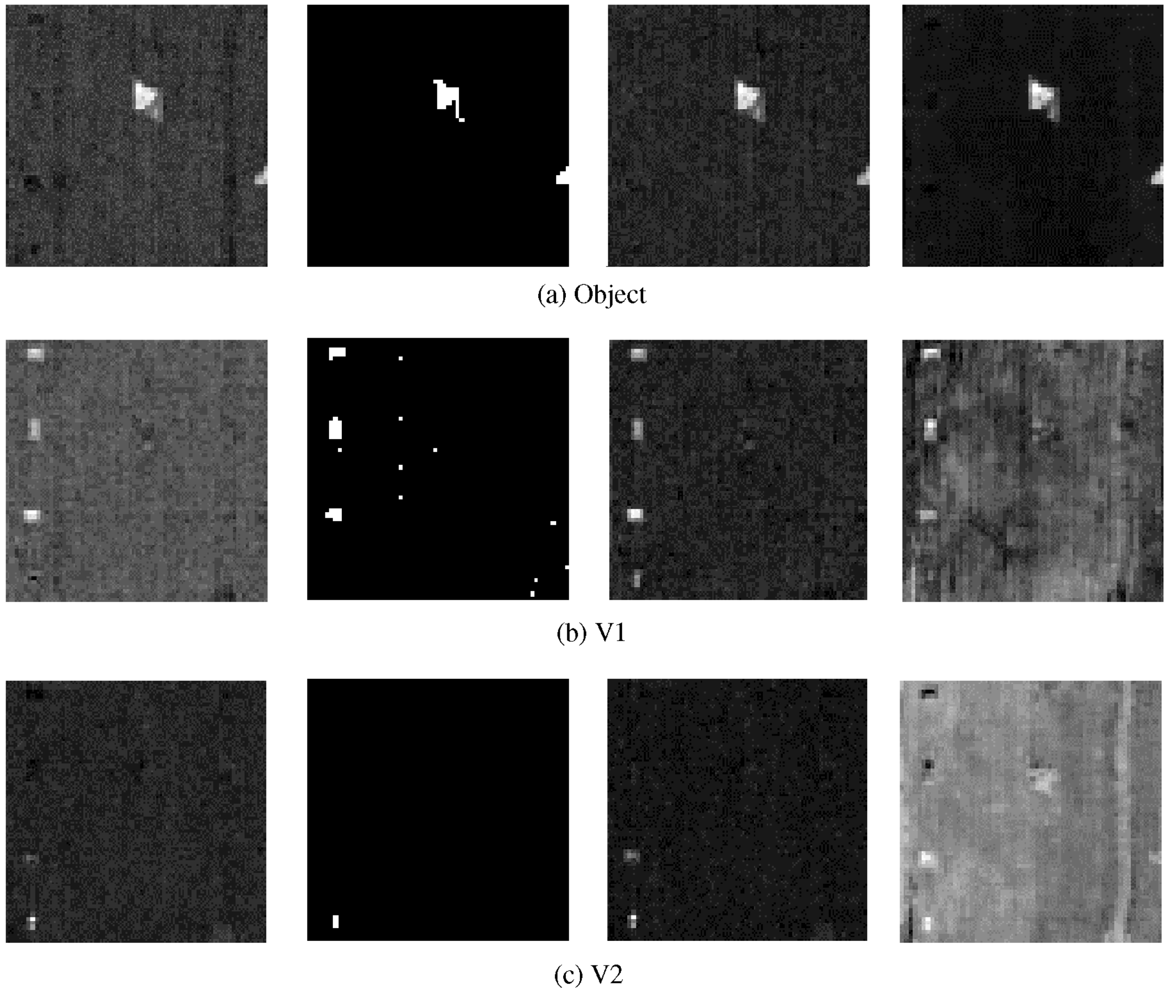
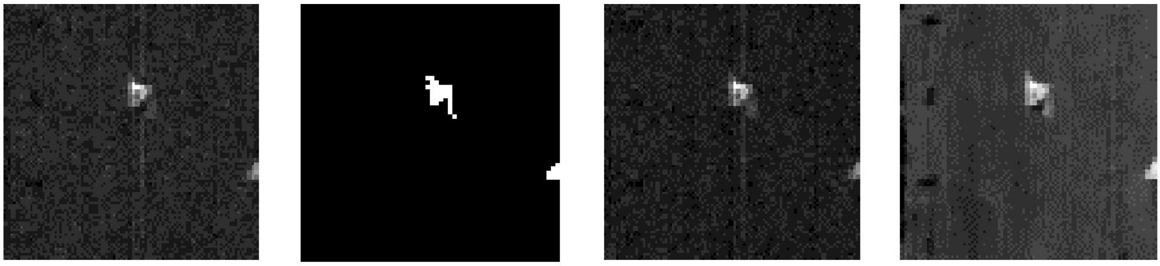


Fig. 5. Results of LCDA, LDA, CEM and OSP. First column: (a) Obj classified by LCDA; (b) V1 classified by LCDA; (c) V2 classified by LCDA. Second column: (a) Obj classified by LDA; (b) V1 classified by LDA; (c) V2 classified by LDA. Third column: (a) Obj classified by CEM; (b) V1 classified by CEM; (c) V2 classified by CEM. Forth column: (a) Obj classified by OSP; (b) V1 classified by OSP; (c) V2 classified by OSP.

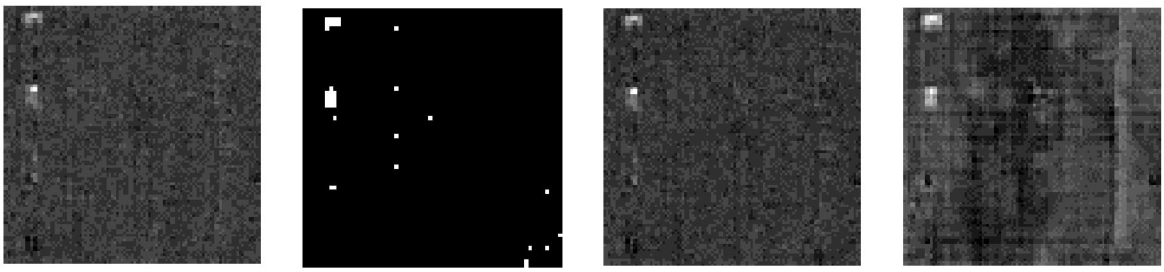
UCEM did as well as ULCDA and its performance was actually improved if V_3 was considered to be a separate type of vehicle. For UOSP it performed well in the sense of target detection, but its performance was slightly offset by extracting small abundance fractions of some background signatures.

7. Conclusion

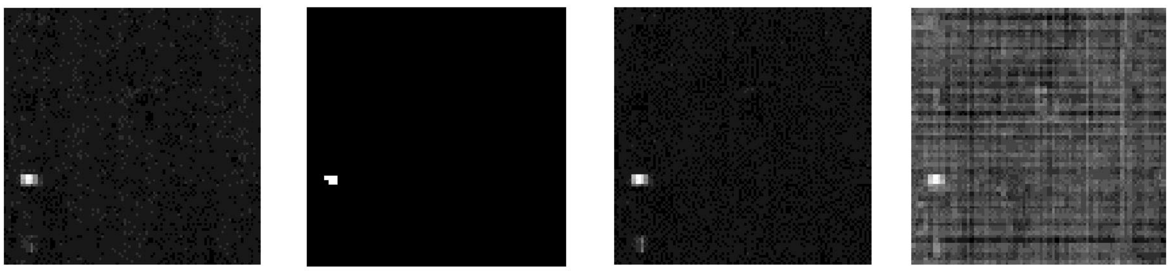
Linear discriminant analysis has been well accepted as a major technique in pattern classification. It can be also applied to hyperspectral image classification [9]. This paper presents a similar but different approach to LDA, called LCDA which replaces Fisher's ratio with the ratio of inter-distance to intra-distance as a criterion for optimality. The advantage of LCDA over LDA is that it constrains the class centers along the desired orthogonal directions. Consequently, all the classes of interest are forced to separate, one must be orthogonal to another. By means of this direction constraint LCDA can detect and classify similar targets. It is particularly useful for very high spatial resolution hyperspectral imagery such as HYDICE [8] where the size of targets ranges from 1 meter to 4 meters. Additionally, LCDA and CEM can be extended to an unsupervised mode for unknown image scenes when no a priori signature knowledge is available. The experimental results are very impressive and almost as good as their supervised counterparts.



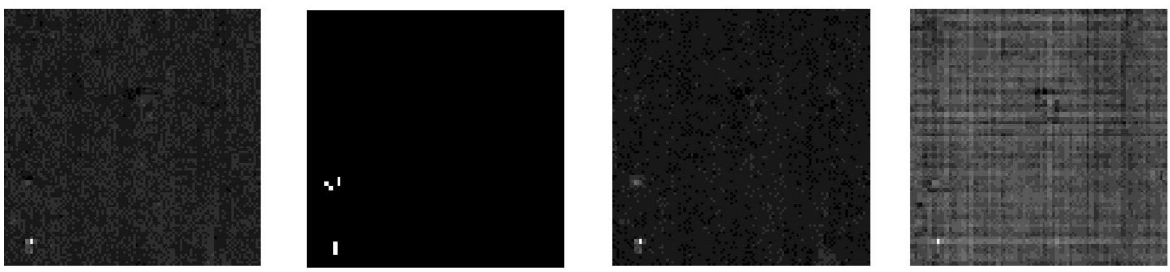
(a) Object



(b) V1(V_1,V_2)



(c) V1(V_3)



(d) V2

Fig. 6. Results of LCDA, LDA, CEM and OSP. First column: (a) Obj classified by LCDA; (b): V1(V_1, V_2) classified by LCDA; (c): V1(V_3) classified by LCDA; (d): V2(V_4) classified by LCDA. Second column: (a) Obj classified by LDA; (b) V1(V_1, V_2) classified by LDA; (c) V1(V_3) classified by LDA; (d): V2(V_4) classified by LDA. Third column: (a) Obj classified by CEM; (b) V1(V_1, V_2) classified by CEM; (c) V1(V_3) classified by CEM; (d): V2(V_4) classified by CEM. Forth column: (a) Obj classified by OSP; (b): V1(V_1, V_2) classified by OSP; (c) V1(V_3) classified by OSP; (d) V2(V_4) classified by OSP.

Acknowledgements

The authors would like to thank Dr. Harsanyi of Applied Signal and Image Technology Inc. for providing AVIRIS data for experiments conducted in this paper.

References

[1] R.A. Schowengerdt, Remote Sensing: Models and Methods for Image Processing, 2nd Edition, Academic Press, New York, 1997.
[2] J.W. Boardman, Inversion of imaging spectrometry data using singular value decomposition, Proceedings of IEEE Symposium. Geoscience and Remote Sensing, 1989, pp. 2069–2072.
[3] J.J. Settle, On the relationship between spectral unmixing and subspace projection, IEEE Trans. Geosci. Remote Sensing 34 (1996) 1045–1046.
[4] Y.E. Shimabukuro, J.A. Smith, The least-squares mixing models to generate fraction images derived from remote sensing multispectral data, IEEE Trans. Geosci. Remote Sensing 29 (1) (1991) 16–20.
[5] T.M. Tu, C.-H. Chen, C.-I. Chang, A least squares orthogonal subspace projection approach to desired signature extraction and detection, IEEE Trans. Geosci. Remote Sensing 35 (1) (1997) 127–139.
[6] C.-I. Chang, X. Zhao, M.L.G. Althouse, J.-J. Pan, Least squares subspace projection approach to mixed pixel classification in hyperspectral images, IEEE Trans. Geosci. Remote Sensing 36 (3) (1998) 898–912.
[7] J.C. Harsanyi, C.-I. Chang, Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection, IEEE Trans. Geosci. Remote Sensing 32 (4) (1994) 779–785.
[8] C.-I. Chang, T.-L.E. Sun, M.L.G. Althouse, An unsupervised interference rejection approach to target detection and classification for hyperspectral imagery, Opt. Engng 37 (3) (1998) 735–743.
[9] C.-I. Chang, H. Ren, An experiment-based quantitative and comparative analysis of hyperspectral target detection and image classification algorithms, IEEE Trans on Geoscience and Remote Sensing (to appear).
[10] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[11] H. Soltanian-Zadeh, P. Windham, D.J. Peck, Optimal linear transformation for MRI feature extraction, IEEE Trans. Med. Imaging 15 (6) (1996) 749–767.
[12] J.T. Tou, R.C. Gonzalez, Pattern Recognition Principles, Addison-Wesley, Reading, MA, 1974.
[13] C.-I. Chang, Q. Du, Interference and noise adjusted principal components analysis, IEEE Trans. Geosci. Remote Sensing 37 (5) (1999) 2387–2396.
[14] B.D. Van Veen, K.M. Buckley, Beamforming: a versatile approach to spatial filtering, IEEE ASSP Mag. (1988) 4–24.
[15] S. Haykin, Adaptive Filter Theory, 3rd Edition, Prentice-Hall, Englewood Cliffs, NJ, 1996.
[16] M.L.G. Althouse, C.-I. Chang, Chemical vapor detection with a multispectral thermal imager, Opt. Engng 30 (11) (1991) 1725–1733.
[17] J.C. Harsanyi, Detection and classification of subpixel spectral signatures in hyperspectral image sequences, Ph.D. Dissertation, Department of Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD, 1993.
[18] J.C. Harsanyi, W. Farrand, C.-I. Chang, Detection of subpixel spectral signatures in hyperspectral image sequences, Annual Meeting, Proceedings of American Society of Photogrammetry & Remote Sensing, Reno, 1994, pp. 236–247.
[19] W. Farrand, J.C. Harsanyi, Mapping the distribution of mine tailing in the coeur d'Alene river valley, Idaho, through the use of constrained energy minimization technique, Remote Sensing Environ. 59 (1997) 64–76.
[20] Q. Du, C.-I. Chang, A linear constrained Euclidean distance-based discriminant analysis for hyperspectral image classification, 1999 Conference on Information and System Science, Johns Hopkins University, Baltimore, MD, 1999.
[21] H. Ren, C.-I. Chang, An unsupervised orthogonal subspace projection to target detection and classification in unknown environment, Spectroradiometric Symposium, San Diego, November 2–7, 1997.
[22] H. Ren, C.-I. Chang, A computer-aided detection and classification method for concealed targets in hyperspectral imagery, International Symposium on Geoscience and Remote Sensing'98, Seattle, WA, July 5–10, 1998, pp. 1016–1018.
[23] H. Ren, C.-I. Chang, A generalized orthogonal subspace projection approach to unsupervised multispectral image classification, Proceedings of SPIE Conference on Image and Signal Processing for Remote Sensing IV, Vol. 3500, Spain, September 21–25, 1998, pp. 42–53.
[24] C.-I. Chang, C. Brumbley, Linear unmixing Kalman filtering approach to signature abundance detection, signature estimation and subpixel classification for remotely sensed images, IEEE Trans. Aerospace Electron. Systems 37 (1) (1999) 319–330.
[25] C.-I. Chang, Spectral information divergence of hyperpsctral image analysis, International Symposium on Geoscience and Remote Sensing'99, June 28-July 2, Hamburg, Germany, 1999, 509–511.

About the Author—QIAN DU received the B.S. and M.S. degrees in electrical engineering from Beijing Institute of Technology in 1992 and 1995, respectively. She is currently a Ph.D. candidate of department of Computer Science and Electrical Engineering, University of Maryland Baltimore County. Her research interests include signal and image processing, pattern recognition and neural networks. Ms. Du is a member of IEEE, SPIE and Phi Kappa Phi.

About the Author—CHEIN-I CHANG received his BS, MS and MA degrees from Soochow University, Taipei, Taiwan, 1973, the Institute of Mathematics at National Tsing Hua University, Hsinchu, Taiwan, 1975 and the State University of New York at Stony

Brook, 1977, respectively, all in mathematics, and MS and MSEE degrees from the University of Illinois at Urbana-Champaign in 1982 respectively and Ph.D. in electrical Engineering from the University of Maryland, College Park in 1987. He was a visiting Assistant Professor from January 1987 to August 1987, Assistant Professor from 1987 to 1993, and is currently an Associate Professor in the Department of Computer Science and Electrical Engineering at the University of Maryland Baltimore County. He was a visiting specialist in the Institute of Information Engineering at the National Cheng Kung University, Tainan, Taiwan from 1994–1995. Dr. Chang is an editor for Journal of High Speed Network and the guest editor of a special issue on Telemedicine and Applications. His research interests include information theory and coding, signal detection and estimation, remote sensing image processing, neural networks, pattern recognition. Dr. Chang is a senior member of IEEE and a member of SPIE, INNS, Phi Kappa Phi and Eta Kappa Nu.